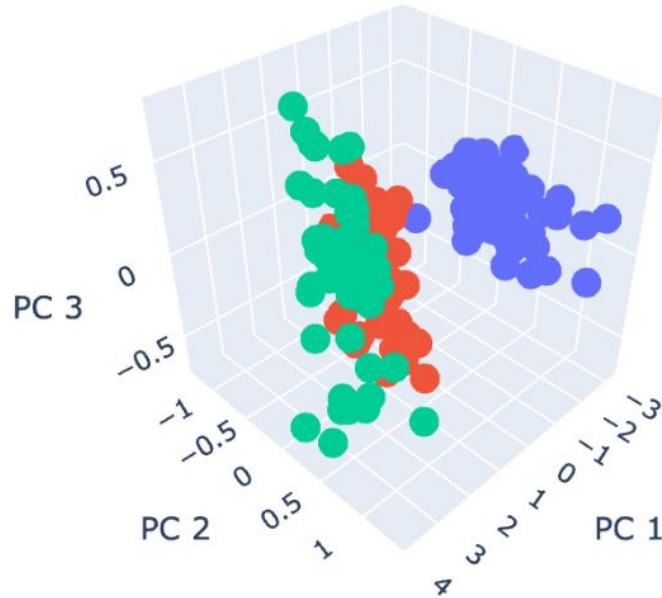


What is Principal Component Analysis?

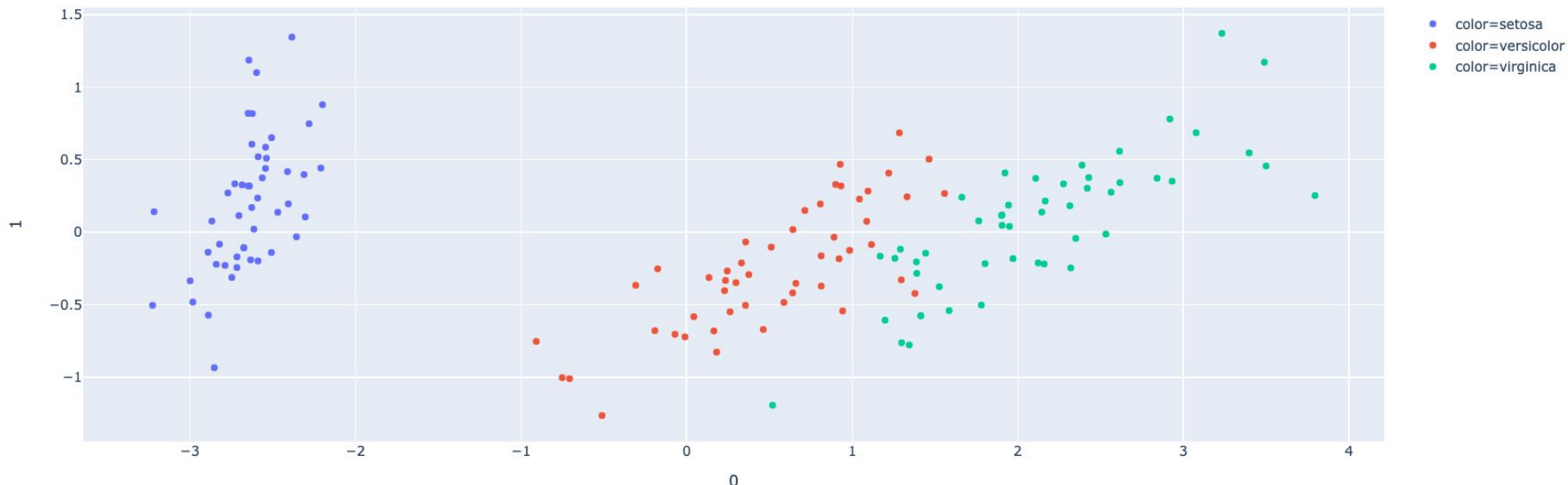


Principal Component Analysis

- Unsupervised
- Non-parametric
- Statistical technique

Primary Use

- **Dimensionality reduction** in machine learning.



High dimensionality

- dataset has a **large number of features**.

High dimensionality **problem**

- **model overfitting**
 - reduces ability to generalize beyond samples in the training set

**As dimensionality of the training dataset
increases...**

- ability to generalize correctly becomes
exponentially harder

High dimensionality

- training set covers a dwindling fraction of the input space

Reduced Feature Set

- **Removes redundant features**
 - boosts learning rates
 - reduces computing costs

Result

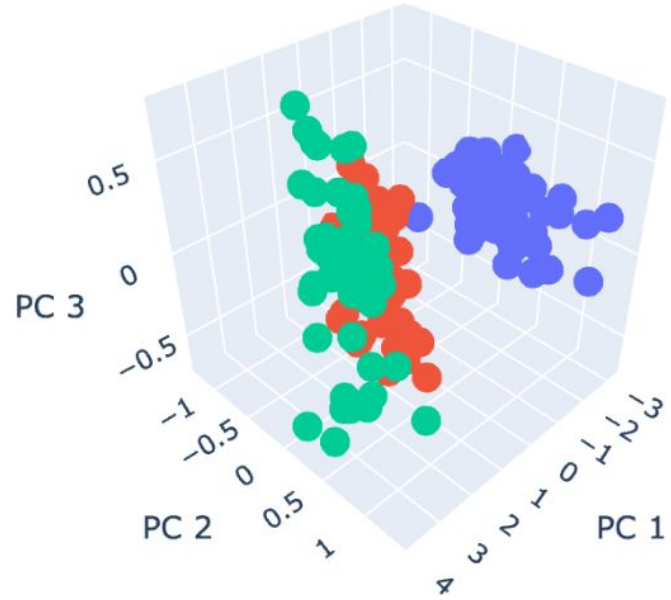
- **Models become more efficient**

Principal Component Analysis

- **can also filter noisy datasets**
 - image compression

1st principal component

- expresses most variance



Each additional component

- less variance and more noise

Principal Component Analysis

- Represents data with a smaller subset of principal components
- Preserves signal
- Discards noise

- makes maximum variability in the dataset more visible by rotating the axes.
- identifies a list of the principal axes to describe dataset
- ranks axes according to variance captured by each.

What is Principal Component Analysis? ✓

